Gaussian Covariance faithful Markov Trees

Dhafer Malouche

Ecole Supérieure de la Statistique
et de l'Analyse de l'Information, Tunisia.
`dhafer.malouche@essai.rnu.tn`

and

Bala Rajaratnam

Standford University, USA.
`brajarat@stanford.edu`

Abstract

A covariance graph is an undirected graph associated with a multivariate probability distribution of a given random vector where each vertex represents each of the different components of the random vector and where the absence of an edge between any pair of variables implies marginal independence between these two variables. Covariance graph models have recently received much attention in the literature and constitute a sub-family of graphical models. Though they are conceptually simple to understand, they are considerably more difficult to analyze. Under some suitable assumption on the probability distribution, covariance graph models can also be used to represent more complex conditional independence relationships between subsets of variables. When the covariance graph captures or reflects all the conditional independence statements present in the probability distribution the latter is said to be faithful to its covariance graph - though no such prior guarantee exists. Despite the increasingly widespread use of these two types of graphical models, to date no deep probabilistic analysis of this class of models, in terms of the faithfulness assumption, is available. Such an analysis is crucial in understanding the ability of the graph, a discrete object, to fully capture the salient features of the probability distribution it aims to describe. In this paper we demonstrate that multivariate Gaussian distributions that have trees as covariance graphs are necessarily faithful. The method of proof is original as it uses an entirely new approach and in the process yields a technique that is novel to the field of graphical models.

1 Introduction

Markov random fields or graphical models are widely used to represent conditional independences in a given multivariate probability distribution (see Kunsch *et al.* (1995), Ji & Seymour (1996), Spitzer (1975), Kindermann & Snell (1980), Lauritzen (1996) to name just a few). Many different types of Markov Random fields or graphical models have been studied in the literature. For example, directed acyclic graphs or DAGs are commonly referred to as "Bayesian networks"

1

(see Pearl (1988)). When the graph is undirected and when such graphs are constructed using marginal independence relationships between pairs of random variables in a given random vector these graphical models are called "covariance graph" models (see Cox & Wermuth (1993), Cox & Wermuth (1996), Kauermann (1996), Malouche & Rajaratnam (2009) and Khare & Rajaratnam (2009)). Covariance graph models are commonly represented by graphs with exclusively bi-directed or dashed edges (see Kauermann (1996)). This representation is used in order to distinguish them from the traditional and widely used concentration graph models. Concentration graphs encode conditional independence between pairs of variables given the remaining ones. Formally, if we consider a random vector $\mathbf{X} = (X_v, v \in V)'$ with a probability distribution P where V is a finite set representing the random variables in $\mathbf{X}$. The concentration graph associated with P is an undirected graph $G = (V, E)$ where

- V is the set of vertices.

- Each vertex represents one variable in $\mathbf{X}$.

- E is the set of edges (between the verices in V) constructed using the pairwise rule : for pair $(u, v) \in V \times V$, $u \neq v$

$$(u, v) \notin E \iff X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{V \setminus \{u,v\}} \tag{1}$$

  where $\mathbf{X}_{V \setminus \{u,v\}} := (X_w, \, w \neq u \text{ and } w \neq v)'$.

Note that $(u, v) \notin E$ means that the vertices u and v are not adjacent in G.

An undirected graph $G_0 = (V, E_0)$ is called the covariance graph associated with the probability distribution P if the set of edges E_0 is constructed as follows

$$(u, v) \notin E \iff X_u \perp\!\!\!\perp X_v \tag{2}$$

The subscript zero is invoked for covariance graphs (i.e., G_0 vs G) as the definition of covariance graphs does not involve conditional independences.

Both concentration and covariance graphs are not only used to encode pairwise relationships between pairs of variables in the random vector $\mathbf{X}$, but as we will see below, these graphs can be used to encode conditional independences that exist between subsets of variables of $\mathbf{X}$. First we introduce some definitions:

The multivariate distribution P is said to satisfy the "intersection property" if for any subsets A, B C and D of V which are pairwise disjoint,

$$\left\{ \begin{array}{l} \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{C \cup D} \\ \text{and} \\ \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_C \mid \mathbf{X}_{B \cup D} \end{array} \right. \quad \text{then} \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B \cup C} \mid \mathbf{X}_D \tag{3}$$

We will call the *intersection* property (see Lauritzen (1996)) in (3) above the *concentration intersection* property in this paper in order to differentiate it from another property that is satisfied by P when studying covariance graph models.

Let P satisfy the concentration intersection property. Then for any triplet (A, B, S) of subsets of V pairwise disjoint, if S separates[1] A and B in the concentration graph G associated with

---

[1]We say that S separates A and B if any path connecting A and B in G intersects S, i.e., $A \perp_G B \mid S$, and is not to be confused with stochastic independence which is denoted by $\perp\!\!\!\perp$ as compared to $\perp_G$.

P then the random vector $\mathbf{X}_A = (X_v, v \in A)'$ is independent of $\mathbf{X}_B = (X_v, v \in B)'$ given $\mathbf{X}_S = (X_v, v \in S)'$. This latter property is called *concentration global Markov* property and is formally defined as,

$$A \perp_G B \mid S \;\Rightarrow\; \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{4}$$

Kauermann (1996) and Banerjee & Richardson (2003) show that if P satisfies the following property : for any triplet (A, B, S) of subsets of V pairwise disjoint,

$$\text{if } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \text{ and } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_C \text{ then } \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B \cup C}, \tag{5}$$

then for any triplet (A, B, S) of subsets of V pairwise disjoint, if $V \setminus (A \cup B \cup S)$ separates A and B in the covariance graph G_0 associated with P then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$. This latter property is called the *covariance global Markov property* and can be written formally as follows

$$A \perp_{G_0} B \mid V \setminus (A \cup B \cup S) \;\Rightarrow\; \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{6}$$

In parallel to the concentration graph case, property (5) will be called the *covariance intersection* property.

Even if P satisfies both intersection properties, the covariance and concentration graphs may not be able to capture or reflect all the conditional independences present in the distribution, i.e., there may exist one or more conditional independences present in the probability distribution that does not correspond to any separation statement in either G or G_0. Equivalently, a lack of a separation statement in the graph does not necessarily imply conditional independences. On the contrary case when no other conditional independence exist in P except the ones encoded by the graph, we classify P as a *faithful* probability distribution to its graphical model. More precisely we say that P is *concentration faithful* to its concentration graph if for any triplet (A, B, S) of subsets of V pairwise disjoint, the following statement holds :

$$S \text{ separates } A \text{ and } B \;\Longleftrightarrow\; \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{7}$$

Similarly, P is said to be *covariance faithful* to its covariance graph G_0 if for any triplet (A, B, S) of subsets of V pairwise disjoint, the following statement holds :

$$V \setminus (A \cup B \cup S) \text{ separates } A \text{ and } B \;\Longleftrightarrow\; \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S. \tag{8}$$

A natural question of both theoretical and applied interest in probability theory is to understand the implications of the *faithfulness* assumption. This assumption is fundamental since it yields a bijection between the probability distribution P and the graph G in terms of the independences that are present in the distribution. In this paper we show that when P is a multivariate Gaussian distribution whose covariance graph are trees are necessarily covariance faithful, i.e., these probability distributions satisfy property (8), i.e., the associated covariance graph G is fully able to capture all the conditional independences present in the multivariate distribution P. This result can be considered as a dual of a previous probabilistic result proved by Becker *et al.* (2005) for concentration graphs that demonstrates that Gaussian distributions having concentration trees, i.e., the concentration graph is a tree are necessarily *concentration faithful* to its concentration graph (implying property (7) is satisfied). This result was proved by showing that Gaussian distributions satisfy an additional intersection property. The approach in the proof of the main

result of this paper is vastly different from the one used for concentration graphs by Becker *et al.* (2005).

The outline of this paper is follows. Section 2 presents graph theory preliminaries. Section 3 gives a brief overview of covariance and concentration graphs associated with multivariate Gaussian distributions. Furthermore, an easier way to encode conditional independence using covariance graphs is given in Section 3. The prove of the main result of this paper is given in Section 4. Section 5 concludes by summarizing the results in the paper and the implications thereof.

2   Graph theory preliminaries

This section introduces notation and terminology that is required in subsequent sections. An undirected graph $G = (V, E)$ consists of two sets V and E, with V representing the set of vertices, and $E \subseteq (V \times V) \setminus \{(u, u), u \in V\}$ the set of edges satisfying :

$$\forall \, (u, v) \in E \iff (v, u) \in E$$

For $u, v \in V$, we write $u \sim_G v$ when $(u, v) \in E$ and we say that u and v are *adjacent* in G.

Definition 1 *A path connecting two distinct vertices u and v in G is a sequence of distinct vertices $(u_0, u_1, \ldots, u_n))$ where $u_0 = u$ and $u_n = v$ where for every $i = 0, \ldots, n-1$, $u_i \sim_G u_{i+1}$.*

Such a path will be denoted $p = p(u, v, G)$ and we say that $p(u, v, G)$ connects u and v or alternatively u and v are *connected* by $p(u, v, G)$. Its *length*, denoted by $|p(u, v, G)|$, is defined as the number of edges connecting the vertices of p. So, in this case $|p(u, v, G)| = n$. We also denote by $\mathcal{P}(u, v, G)$ the set of paths between u and v.

Trees are a particular class of graphs that are studied in this paper. This class of graphs are formally defined below.

Definition 2 *Let $G = (V, E)$ be an undirected graph. The graph G is called a tree if any pair of vertices (u, v) in G are connected by exactly one path, i.e., $|\mathcal{P}(u, v, G)| = 1 \ \forall \, u, v \in V$.*

A subgraph of G *induced* by a subset $U \subseteq V$ is denoted by $G_U = (U, E_U)$, $U \subseteq V$ and $E_U = E \cap (U \times U)$.

Definition 3 *A connected component of a graph G is the largest subgraph $G_U = (U, E_U)$ of G such that each pair of vertices can be connected by at least one path in G_U.*

We now state a Lemma needed in the proof of the main result of this paper.

Lemma 1 *Let $G = (V, E)$ be an undirected graph. If G is a tree, any subgraph of G induced by a subset of V is a union of connected components, each of which are trees (or what we shall refer to as a "union of tree connected components").*

Proof. Consider $U \subset V$, the induced graph G_U and a pair of vertices $(u, v) \in U \times U$. Let us assume to the contrary that u and v are connected by two distinct paths p_1 and p_2 in G_U (i.e., G_U is not a tree). As the set of edges E_U of the graph G_U is included in the set of edges E of G, i.e., $E_U = E \cap (U \times U) \subseteq E$, then p_1 and p_2 are also paths in G. Hence u and v are vertices in G which are connected by two distinct paths, i.e., p_1 and p_2. This of course yields a contradiction with the fact that G is a tree. Thus any pair of vertices in G_U are connected by at most one path and, hence G_U is a union of connected components, each of which are trees (or a "union of tree connected components"). ∎

Definition 4 *For a connected graph, a separator is a subset S of V such that there exists a pair of non-adjacent vertices u and v such that $u, v \notin S$ and*

$$\forall p \in \mathcal{P}(u, v, G), \quad p \cap S \neq \emptyset \tag{9}$$

If S is a separator then it is easily verified that every $S' \supseteq S$ such that $S' \subseteq V \setminus \{u, v\}$ is also a separator. We are thus lead to the notion of a minimal separator.

Definition 5 *The separator S is defined to be a minimal separator between two non-adjacent vertices u and v if for any $w \in S$, the subsets $S \setminus \{w\}$ is not a separator of u and v.*

Note that in the case where G contains more than two connected components and if u and v belong to different connected components the empty set is the only possible separator of u and v. Finally, let A, B and S be pairwise disjoint subsets of V. We say that S *separates* A and B if for any pair of vertices $(u, v) \in A \times B$, any path connecting u and v intersects S. In the case where A and B belong to different connected components of G the subset S can be empty because the set of paths between any pair of vertices $(u, v) \in A \times B$ is empty.

3 Gaussian Concentration and Covariance Graphs

In this section we present a brief overview of concentration and covariance graphs in the case when the probability distribution P is multivariate Gaussian. Such graphical models are commonly referred to as Gaussian covariance or Gaussian concentration graph models.

3.1 Gaussian concentration graph models

Consider a probability space with triplet $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbf{X} : \Omega \to \mathbb{R}^{|V|}$ be a random vector where $\mathbf{X} = (X_v, v \in V)'$ and P represents the induced measure of $\mathbb{P}$ by $\mathbf{X}$. If $\mathbf{X}$ follows a Gaussian distribution then it has the following density function with respect to Lebesgue measure :

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{|V|/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)'\right), \tag{10}$$

where $\mathbf{x} = (x_u, u \in V)' \in \mathbb{R}^{|V|}$, $\mu \in \mathbb{R}^{|V|}$ is the mean vector and $\Sigma = (\sigma_{uv}) \in \mathcal{P}^+$ is the covariance matrix with $\mathcal{P}^+$ denoting the cone of symmetric positive definite matrices. Without loss

of generality we will assume that $\mu = \mathbf{0}$. As any Gaussian distribution with $\mu = \mathbf{0}$ is completely determined by its covariance matrix Σ, this set of multivariate Gaussian distributions can therefore be identified by the set of symmetric positive definite matrices. Gaussian distributions can also be parameterized by the inverse of the covariance matrix Σ denoted by $K = \Sigma^{-1} = (k_{uv})$. The matrix K is called the *precision* or *concentration* matrix. It is well known (see Lauritzen (1996)) that for any pair of variables (X_u, X_v), where $u \neq v$

$$X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{V \setminus \{u,v\}} \iff k_{uv} = 0.$$

Hence the concentration graph $G = (V, E)$ can be constructed simply using the precision matrix K and the following rule

$$(u, v) \notin E \iff k_{uv} = 0.$$

Furthermore it can be easily deduced from a classical result in Hammersly & Clifford (1971), that is reproved in Lauritzen (1996), that any multivariate random vector with a positive density necessarily satisfies the concentration intersection property (3). Hence for Gaussian concentration graph models the pairwise Markov property in (1) is equivalent to the concentration global Markov property in (4).

3.2   Gaussian covariance graph models

As seen earlier in (2) covariance graphs are constructed using pairwise marginal independence relationships. It is also well known that for multivariate Gaussian distributions :

$$X_u \perp\!\!\!\perp X_v \iff \sigma_{uv} = 0.$$

Hence in the Gaussian case the covariance graph $G_0 = (V, E_0)$ can be constructed using the following rule :

$$(u, v) \notin E_0 \iff \sigma_{uv} = 0.$$

It is also easily seen that Gaussian distributions satisfy the covariance intersection property defined in (5). Hence Gaussian covariance graphs can also encode conditional independences according to the following rule : for any triplet (A, B, S) of subsets of V pairwise disjoint, if $V \setminus (A \cup B \cup S)$ separates A and B in the covariance graph G_0 then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$. We now show (see proposition 2 below) that there is a simple way to read conditional independence statements from the covariance graph. This result holds true for any probability distribution that satisfy the covariance intersection property given in (5).

Proposition 2 *Let $\mathbf{X}_V = (X_v, v \in V)'$ be a random vector with probability distribution P satisfying the covariance intersection property in (5) and let $G_0 = (V, E_0)$ be the covariance graph associated with P. Then the following statements are equivalent,*

- *i. for any pairwise disjoint subsets A, B and S of V : if $V \setminus (A \cup B \cup S)$ separates A and B in G_0 then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$*

- *ii. for any pairwise disjoint subsets A, B and S of V : if S separates A and B in G_0 then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V \setminus (A \cup B \cup S)}$*

6

Proof. Let us first assume that (i) is satisfied and let us prove (ii).

Let A, B and S be three pairwise disjoint subsets of V such that S separates A and B in G_0. Note that we can write S as follows:

$$S = V \setminus (V \setminus (A \cup B \cup S)) \cup A \cup B)$$

Since $(V \setminus (A \cup B \cup S) \cup A \cup B = V \setminus S$ and $V \setminus (V \setminus S) = S$.

By hypothesis S separates A and B in G_0. Let $S' = V \setminus (A \cup B \cup S)$ and since $S = V \setminus (S' \cup A \cup B)$ we can apply property (i) to the triplet (A, B, S'). Hence $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{S'}$. Hence $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V \setminus (S \cup A \cup B)}$ since $S' := V \setminus (S \cup A \cup B)$. We have therefore proved that if S separates A and B in G_0, then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V \setminus (S \cup A \cup B)}$.

Assume now that property (ii) is satisfied and let A, B and S be three pairwise disjoint subsets of V such that $V \setminus (S \cup A \cup B)$ separates A and B in G_0. Let us denote by $S' = V \setminus (S \cup A \cup B)$ which is a subset separating A and B in G_0. Since (ii) is satisfied, we deduce that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V \setminus (A \cup B \cup S')}$. However

$$V \setminus (A \cup B \cup S') = V \setminus ((V \setminus (A \cup B \cup S)) \cup A \cup B) = S$$

Hence we conclude that $V \setminus (A \cup B \cup S)$ separates A and B in G_0 implies that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_S$. Thus property (i) is satisfied. ∎

Proposition 2 can be used to formulate an equivalent definition of the covariance faithfulness property.

Definition 6 *Let $\boldsymbol{X}_V = (X_v, \ v \in V)'$ be a random vector with probability distribution P satisfying the covariance intersection property in (5) and let $G_0 = (V, E_0)$ be the covariance graph associated with P. We say that P is covariance faithful to G_0 if for any pairwise disjoint subsets A, B and S of V the following condition is satisfied*

$$S \text{ separates } A \text{ and } B \iff \boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B \mid \boldsymbol{X}_{V \setminus (A \cup B \cup S)}$$

The above reformulation of the covariance faithfulness property is an important ingredient in the proofs in the next section.

4 Gaussian Covariance faithful trees

We now proceed to study the faithfulness assumption in the context of multivariate Gaussian distributions and when the associated covariance graphs are trees.

The main result of this paper, presented in Theorem 3, proves that multivariate Gaussian probability distributions having tree covariance graphs are necessarily faithful to their covariance graphs. The analogous result for concentration graphs was demonstrated by Becker *et al.* (2005) where the authors proved that Gaussian distributions having tree concentration graphs are necessarily faithful to these graphs. We now formally state Theorem 3. The proof follows shortly after a series of lemmas/theorem(s) and an illustrative example.

Theorem 3 *Let $X_V = (X_v, v \in V)'$ be a random vector with Gaussian distribution $P = \mathcal{N}_{|V|}(\mu, \Sigma^{-1})$. Let $G_0 = (V, E_0)$ be the covariance graph associated with P. If G_0 is a tree or more generally a union of connected components each of which are trees (or a union of "tree connected components"), then P is g_0−faithful to G_0.*

The proof of Theorem 3 requires among others a result proved by Jones & West (2005). This result gives a method that can be used to compute the covariance matrix Σ from the precision matrix K using the paths in the concentration graph G. The result can also be easily extended to show that the precision matrix K can be computed from the covariance matrix Σ using the paths in the covariance graph G_0. We now state the result by Jones & West (2005).

Theorem 4 *Jones & West (2005).*
Let $X_V = (X_v, v \in V)'$ be a random vector with Gaussian distribution $P = \mathcal{N}_{|V|}(\mu, \Sigma)$ where Σ and $K = \Sigma^{-1}$ are positive definite matrices. Let $G = (V, E)$ and $G_0 = (V, E_0)$ denote respectively the concentration and covariance graph associated with the probability distribution of X_V.

For all (u, v) in $V \times V$

$$k_{uv} = \sum_{p \in \mathcal{P}(u,v,G_0)} (-1)^{|p|+1} |\sigma|_p \frac{|\Sigma \setminus p|}{|\Sigma|}$$

and

$$\sigma_{uv} = \sum_{p \in \mathcal{P}(u,v,G)} (-1)^{|p|+1} |k|_p \frac{|K \setminus p|}{|K|}$$

where, if $p = (u_0, \ldots, u_n)$,

$$|\sigma|_p = \sigma_{u_0 u_1} \sigma_{u_1 u_2} \ldots \sigma_{u_{n-1} u_n}, \quad |k|_p = k_{u_0 u_1} k_{u_1 u_2} \ldots k_{u_{n-1} u_n},$$

$K \setminus p = (k_{uv}, (u, v) \in (V \setminus p) \times (V \setminus p))$ and $\Sigma \setminus p = (\sigma_{uv}, (u, v) \in (V \setminus p) \times (V \setminus p))$ denote respectively K and Σ with rows and columns corresponding to variables in path p omitted. The determinant of a zero-dimensional matrix is defined to be 1.

The proof of our main theorem (Theorem 3) also requires the results proved in the lemma below.

Lemma 5 *Let $X_V = (X_v, v \in V)'$ be a random vector with Gaussian distribution $P = \mathcal{N}_{|V|}(\mu, K = \Sigma^{-1})$. Let $G_0 = (V, E_0)$ and $G = (V, E)$ denote respectively the covariance and concentration graphs associated with P, then*

   i. *G and G_0 have the same connected components*

   ii. *If a given connected component in G_0 is a tree then the corresponding connected component in G is complete and vice-versa.*

Proof.

Proof of (i).

The fact that G_0 and G have the same connected components can be deduced from the matrix structure of the covariance and the precision matrix. The connected components of G_0 correspond to block diagonal matrices in Σ. Since $K = \Sigma^{-1}$, then by properties of inverting partitioned matrices, K also has the same block diagonal matrices as Σ in terms of the variables that constitute these matrices. These blocks corresponds to distinct components in G and G_0. Hence both matrices have the same connected components.

Proof of (ii).

Let us assume now that the covariance graph G_0 is a tree, hence it is a connected graph with only one connected component. We shall prove that the concentration graph G is complete by using Theorem 4 by Jones & West (2005) and computing any coefficient k_{uv} ($u \neq v$). Since G_0 is a tree, there exists exactly one path between between any two vertices u and v. We shall denote this path as $p = (u_0 = u, \ldots, u_n = v)$. Then by Theorem 4

$$k_{uv} = (-1)^{n+1}\sigma_{u_0 u_1} \ldots \sigma_{u_{n-1} u_n} \frac{|\Sigma \setminus p|}{|\Sigma|} \tag{11}$$

First note that the determinant of the matrices in (11) are all positive since principal minors of positive definite matrices are positive. Second since we are considering a path in G_0, $\sigma_{u_{i-1}u_i} \neq 0, \forall\, i = 1, \ldots, n$. Using these two facts we deduce from (11) that $k_{uv} \neq 0$ for all $(u, v) \in E$. Hence u and v are adjacent in G for all $(u, v) \in E$. The concentration graph G is therefore complete. The proof that when G is assumed to be a tree implying that G_0 is complete follows similarly.

∎

Remark. We further note that Theorem 4 is also directly useful in deducing the completeness of the concentration graph by using the covariance graph in other settings. As a concrete example consider the case when G_0 is a cycle with an even number of edges s.t. $|V| = 2k$ for some odd integer k, and assume that all the coefficients in the covariance matrix Σ of $\mathbf{X}_V$ are positive. Hence a given pair of vertices (u, v) in G_0 are connected by two paths which are both of odd length. Let us denote these paths as p_1 and p_2. Using Theorem 4, it is easily deduced that

$$k_{uv} = \sigma_{|p_1|} \frac{|\Sigma \setminus p_1|}{|\Sigma|} + \sigma_{|p_2|} \frac{|\Sigma \setminus p_2|}{|\Sigma|}$$

Here $|\sigma_{p_1}|$ and $|\sigma_{p_1}|$ are different from zero as they are both equal to a product of positive coefficients. Hence $k_{uv} \neq 0$. The same argument can also be used in the case when p_1 and p_2 both have even length (i.e., $|V| = 2k$ for some even integer k) to deduce that $k_{uv} \neq 0$. Hence u and v are adjacent in the concentration graph G; thus G is necessarily complete.

We now give an example illustrating the main result in this paper (Theorem 3).

Example 1 *Consider a Gaussian random vector $\mathbf{X} = (X_1, \ldots, X_8)'$ with covariance matrix Σ and its associated covariance graph as given in Figure 1. Consider the sets $A = \{1, 2\}$,*
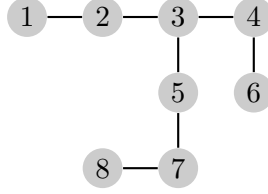
Figure 1: An $8-$vertex covariance tree G_0.

$B = \{5\}$ *and* $S = \{4, 6\}$. *Note that* S *does not separate* A *and* B *in* G_0 *as any path from* A *and* B *does not intersect* S. *In this case we cannot use the covariance global Markov property to claim that* $\boldsymbol{X}_A$ *is not independent of* $\boldsymbol{X}_B$ *given* $\boldsymbol{X}_{V \setminus (A \cup B \cup S)}$. *This is because the covariance global Markov property allows us to read conditional independences present in a distribution if a separation is present in the graph. It is not an "if and only if" property in the sense that the lack of a separation in the graph does not necessarily imply the lack of the corresponding conditional independence. We shall show however that in this example that* $\boldsymbol{X}_A$ *is indeed not independent of* $\boldsymbol{X}_B$ *given* $\boldsymbol{X}_{V \setminus (A \cup B \cup S)}$. *In other words we shall show that the graph has the ability to capture this conditional dependence present in the probability distribution* P.

Let us now examine the relationship between X_2 *and* X_5 *given* $\boldsymbol{X}_{\{3,7,8\}}$. *Note that in this example* $V \setminus (A \cup B \cup S) = \{3, 8, 7\}$, $2 \in A$ *and* $5 \in B$. *Note that the covariance graph associated with the probability distribution of the random vector* $(X_2, X_5, \boldsymbol{X}_{\{3,8,7\}})'$ *is the subgraph represented in Figure 2 and can be obtained directly as a subgraph of* G_0 *induced by the subset* $\{2, 5, 3, 7, 8\}$.
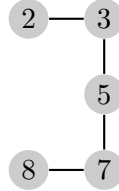

Figure 2: the covariance graph $(G_0)_{\{2,5,3,8,7\}}$

Since 2 *and* 5 *are connected by exactly one path in* $(G_0)_{\{2,5,3,7,8\}}$, *that is* $p = (2, 3, 5)$, *then the coefficient* $k_{25|387}$, *i.e., the coefficient between* 2 *and* 5 *in inverse of the covariance matrix of* $(X_2, X_5, \boldsymbol{X}_{\{3,8,7\}})'$, *can be computed using Theorem 4 as follows*

$$k_{25|387} = (-1)^{2+1} \sigma_{23} \, \sigma_{35} \frac{|\Sigma(\{8, 7\})|}{|\Sigma(\{2, 5, 3, 8, 7\})|} \qquad (12)$$

where $\Sigma(\{7, 8\})$ *and* $\Sigma(\{2, 5, 3, 8, 7\})$ *are respectively the covariance matrices of the Gaussian random vectors* $(X_7, X_8)'$ *and* $(X_2, X_5, \boldsymbol{X}_{\{3,8,7\}})'$. *Hence* $k_{25|387} \neq 0$ *since the right hand side of the equation in (12) is different from zero. Hence* $X_2 \not\perp\!\!\!\perp X_5 \mid \boldsymbol{X}_{\{3,8,7\}}$.

Now recall that for any Gaussian random vector vector $\boldsymbol{X}_V = (X_u, \ u \in V)'$,

$$\boldsymbol{X}_A \perp\!\!\!\perp \boldsymbol{X}_B \mid \boldsymbol{X}_C \text{ if and only if } \forall (u, v) \in A \times B, \ X_u \perp\!\!\!\perp X_v \mid \boldsymbol{X}_C \qquad (13)$$

where A, B and C are pairwise disjoint subsets of V. The contrapositive of (13) yields

$$X_2 \not\perp\!\!\!\perp X_5 \mid \boldsymbol{X}_{\{3,7,8\}} \;\Rightarrow\; \boldsymbol{X}_{\{1,2\}} \not\perp\!\!\!\perp X_5 \mid \boldsymbol{X}_{\{3,7,8\}}.$$

Hence we conclude that since $\{3,7,8\}$ does not separate $\{1,2\}$ and $\{5\}$ therefore $\boldsymbol{X}_{\{1,2\}}$ is not independent of X_5 given $\boldsymbol{X}_{\{3,7,8\}}$, i.e.,

$$\{1,2\} \not\perp_{G_0} \{5\} \mid \{3,7,8\} \Rightarrow \boldsymbol{X}_{\{1,2\}} \not\perp\!\!\!\perp X_5 \mid \boldsymbol{X}_{\{3,7,8\}}$$

.

We now proceed to the proof of Theorem 3. **Proof.** of Theorem 3. Without loss of generality we assume that G_0 is a connected tree. Let us assume to the contrary that P is not covariance faithful to G_0, then there exists a triplet (A, B, S) of pairwise disjoint subsets of V, such that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V\setminus(A\cup B\cup S)}$, but S does not separate A and B in G_0, i.e.,

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V\setminus(A\cup B\cup S)} \;\text{ and }\; A \not\perp_{G_0} B \mid S$$

As S does not separate A and B and since G_0 is a connected tree, then there exists a pair of vertices $(u, v) \in A \times B$ such that the single path p connecting u and v in G_0 does not intersect S, i.e., $S \cap p = \emptyset$. Hence $p \subseteq V \setminus S = (A \cup B) \cup (V \setminus (A \cup B \cup S))$. Thus two cases are possible with regards to where the path p can lie : either $p \subseteq A \cup B$ or $p \cap (V \setminus (A \cup B \cup S)) \neq \emptyset$. Let us examine both cases separately.

- **Case 1 :** $p \subseteq A \cup B$

  In this case the entire path between u and v lies in $A \cup B$ and hence we can find a pair of vertices[2] (u', v') belonging to p and $(u', v') \in A \times B$ such that $u' \sim_{G_0} v'$.

  Recall that since G_0 is a tree, any induced graph of G_0 by a subset of V is a union of tree connected components (see Lemma 1). Hence the subgraph $(G_0)_W$ of G_0 induced by $W = \{u', v'\} \cup V \setminus (A \cup B \cup S)$ is a union of tree connected components. As u' and v' are adjacent in G_0, they are also adjacent in $(G_0)_W$ and belong to the same connected component[3] of $(G_0)_W$. Hence the only path between u' and v' is precisely the edge (u', v'). Using theorem 4 to compute the coefficient $k_{u'v'|V\setminus(A\cup B\cup S)}$, i.e., $(u', v')th$ coefficient in the inverse of the covariance matrix of the random vector $\mathbf{X}_W = (X_w, w \in W)' = (X_{u'}, X_{v'}, \mathbf{X}_{V\setminus(A\cup B\cup S)})'$, we obtain,

  $$k_{u'v'|V\setminus(A\cup B\cup S)} = (-1)^{1+1}\sigma_{u'v'}\frac{|\Sigma(W \setminus \{u', v'\})|}{|\Sigma(W)|}, \tag{14}$$

  where $\Sigma(W)$ denotes the covariance matrix of $\mathbf{X}_W$, and $\Sigma(W\setminus\{u', v'\})$ denotes the matrix $\Sigma(W)$ with the rows and the columns corresponding to variables $X_{u'}$ and $X_{v'}$ omitted. We

---

[2]As an illustration of this point consider the graph presented in Figure 1. Let $A = \{1, 2\}$, $B = \{3, 5\}$ and $S = \{4, 6\}$. We note that the path $p = (1, 2, 3, 5)$ lies entirely in $A \cup B$ and hence we can find two vertices, namely, $2 \in A$ and $3 \in B$, belonging to path p that are adjacent in G_0.

[3]In our example in Figure 1 with $W = \{2, 3, 8, 7\}$, $(G_0)_W$ consists a union of two connected components with its respective vertices being $\{2, 3\}$ and $\{8, 7\}$.

can therefore deduce from (14) that $k_{u'v'|V\backslash(A\cup B\cup S)} \neq 0$. Recall that at the start of the proof we assumed to the contrary that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V\backslash(A\cup B\cup S)}$. Now since P is Gaussian, for pairwise disjoint subsets $A, B, V \backslash (A \cup B \cup C)$ then

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{V\backslash(A\cup B\cup C)} \Leftrightarrow \forall\, (u,v) \in A \times B,\ X_u \perp\!\!\!\perp X_v \mid \mathbf{X}_{V\backslash(A\cup B\cup C)} \qquad (15)$$

Note however that we have established that $X_{u'} \not\perp\!\!\!\perp X_{v'} \mid \mathbf{X}_{V\backslash(A\cup B\cup S)}$ since $k_{u'v'|V\backslash(A\cup B\cup S)} \neq 0$. Hence we obtain a contradiction to (15) since $u' \in A$ and $v' \in B$.

- **Case 2 :** $p \cap (V \backslash (A \cup B \cup S)) \neq \emptyset$ & $V \backslash (A \cup B \cup S)$ is not empty.

  Now if $V \backslash (A \cup B \cup S)$ is empty then p has to lie entirely in $A \cup B$. This is because by assumption p does not intersect S. The case when p lies in $A \cup B$ is covered in Case 1 and hence it is assumed that $V \backslash (A \cup B \cup S) \neq \emptyset$. [4]

  In this case there exists a pair of vertices $(u', v') \in A \times B$ with $u', v' \in p$, such that the vertices u' and v' are connected by exactly one path $p' \subseteq p$ in the induced graph $(G_0)_W$ of G_0 by $W = \{u', v'\} \cup V \backslash (A \cup B \cup S)$ (see Lemma 1) [5].

  Let us now use Theorem 4 to compute the coefficient $k_{u'v'|V\backslash(A\cup B\cup S)}$, i.e., the $(u', v')-$coefficient in the inverse of the covariance matrix of the random vector $\mathbf{X}_W = (X_w,\ w \in W)' = (X_{u'}, X_{v'}, \mathbf{X}_{V\backslash(A\cup B\cup S)})'$. We obtain that

  $$k_{u'v'|V\backslash(A\cup B\cup S)} = (-1)^{|p'|+1}|\sigma_{p'}|\,\frac{|\Sigma(W \backslash p')|}{|\Sigma(W)|}, \qquad (16)$$

  where $\Sigma(W)$ denotes the covariance matrix of $\mathbf{X}_W$ and $\Sigma(W \backslash p')$ denotes $\Sigma(W)$ with the rows and the columns corresponding to variables in path p' omitted. One can therefore easily deduce from (16) that $k_{u'v'|V\backslash(A\cup B\cup S)} \neq 0$. Thus $X_{u'}$ is not independent of $X_{v'}$ given $\mathbf{X}_{V\backslash(A\cup B\cup S)}$. Hence once more we obtain a contradiction to (15) since $u' \in A$ and $v' \in B$.

∎

Remark. The dual result of the theorem above for the case of concentration trees was proved by Becker *et al.* (2005). We note however that the argument used in the proof of Theorem 3 cannot also be used to prove faithfulness of Gaussian distributions that have trees as concentration graphs. The reason for this is as follows. In our proof we employed the fact that the sub-graph $(G_0)_{\{u,v\}\cup S}$ of G_0 induced by a subset $\{u,v\} \cup S \subseteq V$ is also the covariance graph associated with the Gaussian sub-random vector of $\mathbf{X}_V$ as denoted by $\mathbf{X}_{\{u,v\}\cup S} = (X_w,\ w \in \{u,v\} \cup S)'$. Hence it was possible to compute the coefficient $k_{uv|S}$ which quantifies the conditional (in)dependence between u and v given S, in terms of the paths in $(G_0)_{\{u,v\}\cup S}$

---

[4] As an illustration of this point consider once more the graph presented in Figure 1. Consider $A = \{1, 2\}$, $B = \{7, 8\}$ and $S = \{4, 6\}$. Here $V \backslash (A \cup B \cup S) = \{3, 5\}$ and the path $p = (1, 2, 3, 5, 7, 8)$ connecting A and B intersects $V \backslash (A \cup B \cup S)$.

[5] In our example in figure 1 with $A = \{1, 2\}$, $B = \{7, 8\}$ and $S = \{4, 6\}$, the vertices u' and v' will correspond to vertices 2 and 7 respectively, and $p' = (2, 3, 5, 7)$, which is a path entirely contained in $V \backslash (A \cup B \cup S) \cup \{u', v'\}$.

and the coefficients of the covariance matrix of $\mathbf{X}_{\{u,v\}\cup S} = (X_w, u \in \{u,v\} \cup S)'$. On the contrary, in the case of concentration graphs the sub-graph $G_{\{u,v\}\cup S}$ of the concentration graph G induced by $\{u,v\} \cup S$ is not in general the concentration graph of the random vector $\mathbf{X}_{\{u,v\}\cup S} = (X_w, u \in \{u,v\} \cup S)'$. Hence our approach is not directly applicable in the concentration graph setting.

5   Conclusion

Faithfulness of a probability distribution to a graph is a crucial assumption that is often made in the probabilistic treatment of graphical models. This assumption describes the ability of a graph to reflect or encode the multivariate dependencies that are present in a joint probability distribution. Much of the methodology in this area often do not undertake a detailed analysis of the faithfulness assumption, as such an endeavor requires a more careful and rigorous probabilistic study of the joint distribution at hand. In this note we looked at the class of multivariate Gaussian distributions that are Markov with respect to covariance graphs and prove that Gaussian distributions which have trees as their covariance graphs are necessarily faithful. The method of proof that is employed in this paper is novel in the sense that it is self contained and yields a completely new approach to demonstrating faithfulness - as compared to the methods that are traditionally used in the literature. Moreover, it is also vastly different in nature from the proof of the analogous result for concentration graph models. Hence the approach used in this paper promises to have further implications and give other insights. Future research in this area will explore if the techniques used in this paper can be modified to prove or disprove faithfulness for other classes of graphs.

Acknowledgments

References

Banerjee, M., & Richardson, T. 2003. On a Dualization of Graphical Gaussian Models: A Correction Note. *Scand. J. Statist.*, **Vol 30**, 817–820.

Becker, Ann, Geiger, Dan, & Meek, Christopher. 2005. Perfect Tree-like Markovian Distributions. *Probability and Mathematical Statistics*, **25**(2), 231–239.

Cox, D. R., & Wermuth, N. 1996. *Multivariate Depencies : Models, Analysis and Interpretations*. Chapman and Hall.

Cox, D.R., & Wermuth, M. 1993. Linear dependencies represented by chain graphs (with Discussion). *Statist. Sci.*, **8**, 204–218, 247–277.

Hammersly, J. M., & Clifford, P. E. 1971. *Markov fields on finite graphs and lattices*. Unpublished manuscript.

Ji, C., & Seymour, L. 1996. A consistent model selection procedure for Markov random fields based on penalized pseudolikelihood. *The Annals of Applied Probability*, **6**(2), 423–443.

Jones, B., & West, M. 2005. Covariance decomposition in undirected Gaussian graphical models. *Biometrika*, **92**, 770–786.

Kauermann, G. 1996. On a dualization of graphical Gaussian models. *Scand. J. Statist.*, **23**, 105–116.

Khare, K., & Rajaratnam, B. 2009. Wishart distributions for decomposable covariance graph models. *under review in the Annals of Statistics.*

Kindermann, R., & Snell, J. L. 1980. *Markov Random Fields and Their Applications*. American Mathematical Society, Providence, Rhode Island.

Kunsch, H., Gemanand, S., & Kehagias, A. 1995. Hidden Markov Random Fields. *The Annals of Applied Probability*, **5**(3), 577–602.

Lauritzen, S. L. 1996. *Graphical Models*. New York : Oxford University Press.

Malouche, D., & Rajaratnam, B. 2009. Analysis of the faithfulness assumption in Graphical Models. *Technical Report, Department of Statistics, Stanford University.*

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Tech. rept. Morgan Kaufman.

Spitzer, C. 1975. Markov random fields on an infinite tree. *The Annals of Probability*, **3**, 387–398.